INFORMATION AND CONTROL 35, 1–19 (1977)

On Sliding Block Coding for Transmission of a Source over a Stationary Nonanticipatory Channel

JOHN C. KIEFFER

Department of Mathematics, University of Missouri, Rolla, Missouri 65401

It is shown that the capacity C_b of a stationary nonanticipatory channel with respect to block coding is at least as great as the capacity C_s of the channel with respect to sliding block coding. For several types of stationary nonanticipatory channels (namely, ergodic channels, channels with additive random noise, and averaged channels whose components are discrete memoryless channels), it is shown that $C_b = C_s$, thereby generalizing a result of Gray and Ornstein for the discrete memoryless channel.

I. INTRODUCTION

Notation

Let S be a set. We use $|S|$ to denote the cardinality of S. Z denotes the set of integers. S^∞ denotes the set of all doubly infinite sequences $(s_i)_{i=-\infty}^{\infty}$ from S. For $n = 1, 2,\ldots$, let $S^n = \{(s_i)_1^n: s_i \in S,\ i = 1, 2,\ldots, n\}$.
If $s = (s_i) \in S^\infty$ or S^n and j, k are integers with $j \leqslant k$ such that $\{i: j \leqslant i \leqslant k\}$ is contained in the domain of s, let s_j^k (or $[s]_j^k$) denote the sequence $(s_i')_1^{k-j+1}$ in S^{k-j+1} such that $s_i' = s_{i+j-1}$, $i = 1,\ldots, k - j + 1$. But, to represent the element of the sequence $s \in S^\infty$ or S^n corresponding to the integer i, we write s_i (or $[s]_i$) rather than s_i^i.

If $W = \{W_i\}_{i=-\infty}^{\infty}$ is a sequence of S-valued functions whose domain is some set Ω, and $n \in Z$, let W_n^∞ denote the sequence $W_n^\infty = \{W_i\}_{i=n}^{\infty}$. If m, n are integers with $m \leqslant n$, let W_m^n denote the finite sequence of functions $W_m^n = \{W_i\}_{i=m}^{n}$. We regard W_m^n as the S^{n-m+1}-valued function such that $W_m^n(\omega) = (s_i)_1^{n-m+1}$ where $s_i = W_{i+m-1}(\omega)$, $\omega \in \Omega$.

$\mathscr{S}^\infty$ will denote the usual product σ-field of subsets of S^∞. T_S: $S \to S$ will denote the two-sided shift on S^∞; that is, if $s = (s_i) \in S^\infty$, then $T_S(s) = s' = (s_i')$, where $s_i' = s_{i+1}$. Let $\mathscr{M}_S$ be the sub-σ-field of $\mathscr{S}^\infty$ consisting of the T_S-invariant sets.

Sources

By a source $[A, P]$ we mean a finite alphabet A and a probability measure P on $\mathscr{A}^\infty$ stationary with respect to T_A. A sequence $u \in A^\infty$ is periodic if $T_A{}^N(u) = u$

ISSN 0019-9958

for some positive integer N. The smallest positive integer N such that $T_A{}^N(u) = u$ is called the period of u. A source $[A, P]$ is aperiodic if P assigns probability zero to the (countable) set of periodic points. A source $[A, P]$ is periodic if P assigns probability one to the set of periodic points. A source $[A, P]$ is ergodic if P is ergodic with respect to T_A. (We remark that every ergodic source is either periodic or aperiodic.) A source $[A, P]$ is iid if P is a product measure on $\mathscr{A}^\infty$. Define $H(P)$ to be the entropy of the source $[A, P]$.

Channels

In this paper a channel will always mean a stationary nonanticipatory channel; that is, a triple $[B, C, \mu]$ where B, C are finite alphabets and $\mu(\cdot \mid \cdot)$: $\mathscr{C}^\infty \times B^\infty \to [0, 1]$ is the channel probability function, which we assume satisfies the following conditions:

1. $\mu(\cdot \mid x)$ is a probability measure on $\mathscr{C}^\infty$, $x \in B^\infty$.
2. $\mu(E \mid \cdot)$ is $\mathscr{B}^\infty$-measurable, $E \in \mathscr{C}^\infty$.
3. $\mu(T_C E \mid T_B x) = \mu(E \mid x)$, $x \in B^\infty$, $E \in \mathscr{C}^\infty$.
4. For $N = 1, 2,\ldots,$ if $E \subset C^N$ and $x, x' \in B^\infty$ satisfy $[x]_1^N = [x']_1^N$, then $\mu(\{y \in C^\infty : y_1{}^N \in E\} \mid x) = \mu(\{y \in C^\infty : y_1{}^N \in E\} \mid x')$.

If $[B, C, \mu]$ is a channel, then for $n = 1, 2,\ldots,$ let $\mu_n(\cdot \mid \cdot)$: $C^n \times B^n \to [0, 1]$ be the conditional probability mass function (PMF) such that $\mu_n(y' \mid x') = \mu(\{y \in C^\infty : y_1{}^n = y'\} \mid x)$, where $y' \in C^n$, $x' \in B^n$, and $x \in B^\infty$ satizfies $x_1{}^n = x'$. If $E \subset C^n$, and $x \in B^n$, define $\mu_n(E \mid x) = \sum_{y \in C^n} \mu_n(y \mid x)$.

Given the channel $[B, C, \mu]$, the sequence $\{\mu_n\}_1^\infty$ of conditional PMF's satisfies:

1. For $n = 1, 2,\ldots, \sum_{y \in C} \mu_{n+1}(yy' \mid xx') = \mu_n(y' \mid x'), y' \in C^n$, $x' \in B^n$, $x \in B$.
2. For $n = 1, 2,\ldots,$ $\sum_{y \in C} \mu_{n+1}(y'y \mid x'x) = \mu_n(y' \mid x')$, $y' \in C^n$, $x' \in B^n$, $x \in B$.

Conversely, given any sequence of conditional PMF's $\mu_n(\cdot \mid \cdot)$: $B^n \times C^n \to [0, 1]$, $n = 1, 2,\ldots,$ which satisfy the two conditions above, then there is a unique channel $[B, C, \mu]$ which induces the given sequence $\{\mu_n\}$.

A (n, N, λ) code for the channel $[B, C, \mu]$ consists of N points $w_1, w_2,\ldots, w_N$ in B^n, and N pairwise disjoint subsets $E_1,\ldots, E_N$ of C^n such that $\mu_n(E_i \mid w_i) > 1 - \lambda$, $i = 1, 2,\ldots, N$.

If $[B, P]$ is a source and $[B, C, \mu]$ is a channel, let $P\mu$ denote the probability measure on the product σ-field $\mathscr{B}^\infty \times \mathscr{C}^\infty$ such that $P\mu(E \times F) = \int_E \mu(F \mid x) dP(x)$.

The Source–Channel Hookup

Let $[A, P]$ be a source which we want to transmit over the channel $[B, C, \mu]$. The source chooses a random $u \in A^\infty$, which is coded into a sequence $x \in B^\infty$.

Then x is transmitted over the channel, yielding a random $y \in C^\infty$. This $y \in C^\infty$ is decoded into a $v \in A^\infty$, which we hope will be u. To analyze the performance of our coding and decoding schemes, the measurable space $(A^\infty \times B^\infty \times C^\infty \times A^\infty, \mathscr{A}^\infty \times \mathscr{B}^\infty \times \mathscr{C}^\infty \times \mathscr{A}^\infty)$ will prove useful. On this measurable space, we define sequences $U = \{U_i\}_{-\infty}^\infty$, $X = \{X_i\}_{-\infty}^\infty$, $Y = \{Y_i\}_{-\infty}^\infty$, $V = \{V_i\}_{-\infty}^\infty$ of, respectively, A-valued, B-valued, C-valued, and A-valued measurable functions such that $U_i(u, x, y, v) = u_i$, $X_i(u, x, y, v) = x_i$, $Y_i(u, x, y, v) = y_i$, $V_i(u, x, y, v) = v_i$, $u \in A^\infty$, $x \in B^\infty$, $y \in C^\infty$, $v \in A^\infty$, $i \in Z$. A coder is a measurable map ϕ: $A^\infty \to B^\infty$. A decoder is a measurable map ψ: $C^\infty \to A^\infty$. Transmitting the source $[A, P]$ over the channel $[B, C, \mu]$ using the coder ϕ and the decoder ψ yields a probability measure $P^{\phi,\psi}$ on $\mathscr{A}^\infty \times \mathscr{B}^\infty \times \mathscr{C}^\infty \times \mathscr{A}^\infty$ such that

$$P^{\phi,\psi}(E \times F \times G \times H) = \int_{\phi^{-1}(F) \cap E} \mu(\psi^{-1}(H) \cap G \mid \phi(u))\, dP(u),$$

$$E \in \mathscr{A}^\infty, \quad F \in \mathscr{B}^\infty, \quad G \in \mathscr{C}^\infty, \quad H \in \mathscr{A}^\infty. \qquad (1)$$

An analysis of $P^{\phi,\psi}$ will give us the reliability of the coder and decoder in reproducing the transmitted source.

Block Coding

Let N be a positive integer. A coder ϕ is a block coder of type N if there exists a map ϕ_N: $A^N \to B^N$ such that $\phi(u)_{kN+1}^{kN+N} = \phi_N(u_{kN+1}^{kN+N})$, $k \in Z$. A coder ϕ is called a block coder if it is a block coder of type N for some N. Similarly, we define a block decoder of type N and a block decoder. We say the source $[A, P]$ is transmissible over the channel $[B, C, \mu]$ with respect to block coding if for any $\epsilon > 0$, there exists a positive integer N, a block coder ϕ of type N and a block decoder ψ of type N such that $P^{\phi,\psi}[U_1^N = V_1^N] > 1 - \epsilon$. We define the block coding capacity C_b of the channel $[B, C, \mu]$ to be $C_b = \sup_{[A,P]} H(P)$, where the supremum is taken over all ergodic sources $[A, P]$ transmissible over the channel with respect to block coding, and all finite alphabets A. (There is always at least one such source; investigate the case where A consists of a single point.)

We have the following coding theorem and converse for block coding, which appears in Winkelbauer (1960).

THEOREM 1. (a) *Let $[A, P]$ be an ergodic source such that $H(P) < C_b$. Then $[A, P]$ is transmissible over the channel $[B, C, \mu]$ with respect to block coding. Moreover, for any $\epsilon > 0$, there exists for sufficiently large N a block coder ϕ of type N and a block decoder ψ of type N such that $P^{\phi,\psi}[U_1^N = V_1^N] > 1 - \epsilon$.*

(b) *If $H(P) > C_b$ then $[A, P]$ is not transmissible over $[B, C, \mu]$ with respect to block coding.*

Recently, workers in information theory have been trying to prove results of this type using a new type of coding due to Gray (1975), called sliding block coding.

Sliding Block Coding

A coder $\phi: A^\infty \to B^\infty$ is called an invariant coder if $\phi \cdot T_A = T_B \cdot \phi$. A coder ϕ is called a sliding block coder of type N if there exists a map $\phi_N: A^{2N+1} \to B$ such that $\phi(u)_i = \phi_N(u_{i-N}^{i+N})$, $i \in Z$. A coder ϕ is called a sliding block coder if it is a sliding block coder of type N for some N. Similarly, one defines invariant decoders, sliding block decoders of type N, and sliding block decoders. A source $[A, P]$ is said to be transmissible over the channel $[B, C, \mu]$ with respect to sliding block coding if for any $\epsilon > 0$, there exists a sliding block coder ϕ and a sliding block decoder ψ such that $P^{\phi,\psi}[U_0 = V_0] > 1 - \epsilon$. We define the sliding block capacity C_s of the channel $[B, C, \mu]$ to be $\sup_{[A,P]} H(P)$, where the supremum is over all ergodic sources $[A, P]$ transmissible over the channel with respect to sliding block coding, and all finite alphabets A. (As pointed out earlier with block coding, there is always at least one such source.) There are two natural conjectures which present themselves. They were proved in Gray and Ornstein (1976) for the special case of the DMC (discrete memoryless channel).

Conjecture 1. $C_s = C_b$ for every channel $[B, C, \mu]$.

Conjecture 2. (a) If $[A, P]$ is an aperiodic ergodic source with $H(P) < C_s$, then $[A, P]$ is transmissible over $[B, C, \mu]$ with respect to sliding block coding.

(b) If $[A, P]$ is an ergodic source satisfying $H(P) > C_s$, then $[A, P]$ is not transmissible over $[B, C, \mu]$ with respect to sliding block coding.

(Of course, Conjecture 2(b) is obviously true from the way in which C_s was defined.) We were not able to settle these conjectures in general. It would be nice to know whether $C_s = C_b$ holds in general, because if so, the general formula given in Kieffer (1974a) for C_b could be used to calculate C_s as well. In this paper we show that $C_s \leqslant C_b$ holds in general and the Conjectures 1 and 2 both hold if $[B, C, \mu]$ is either an ergodic channel, or a channel with additive random noise, or an averaged channel whose components are DMC's. We also show that Conjecture 2(a) fails in general for periodic ergodic sources.

II. A Demonstration that $C_s \leqslant C_b$

THEOREM 2. *For any channel* $[B, C, \mu]$, $C_s \leqslant C_b$ *holds.*

Proof. The proof is an adaptation of an idea appearing in Nedoma (1957, pp. 159–162). Suppose $C_s > C_b$. Then there exists a finite alphabet A with $|A| \geqslant 2$ and an ergodic source $[A, P]$ such that $H(P) > C_b$ and $[A, P]$ is transmissible over the channel with respect to sliding block coding. We show that this leads to a contradiction. Fix ϵ, $0 < \epsilon < \frac{1}{4}$. Then there exists a sliding block coder ϕ of type M and a sliding block decoder ψ of type N such that $P^{\phi,\psi}[U_0 \neq V_0] < \epsilon^4$. Let $d: A \times A \to [0, 1]$ be the Hamming metric. (That is,

$d(u, v) = 1$ if $u \neq v$; $d(u, v) = 0$ if $u = v$.) For $n = 1, 2, \ldots$ define a distortion metric d_n: $A^{2n+1} \times A^{2n+1} \to [0, 1]$ such that

$$d_n(u, v) = (2n+1)^{-1} \sum_{i=1}^{2n+1} d(u_i, v_i), \qquad u, v \in A^{2n+1}.$$

Since $\{(U_i, V_i)\}$ is stationary with respect to $P^{\phi,\psi}$, we have $E[d_n(U_{-n}^n, V_{-n}^n)] = E[d(U_0, V_0)] = P^{\phi,\psi}[U_0 \neq V_0] < \epsilon^4$. For each n, if $y \in C^\infty$, the coordinates $-n$ through n of $\psi(y)$ depend only on the coordinates $-n-N$ through $n+N$ of y. Hence, there is a map ψ_n: $C^{2n+2N+1} \to A^{2n+1}$ such that for $y \in C^\infty$, $\psi(y)_{-n}^n = \psi_n(y_{-n-N}^{n+N})$. Similarly, for each n, there is a map ϕ_n: $A^{2n+2M+2N+1} \to B^{2n+2N+1}$ such that if $u \in A^\infty$ then $\phi(u)_{-n-N}^{n+N} = \phi_n(u_{-n-M-N}^{n+M+N})$. Also, for each n, let π_n: $A^{2n+2M+2N+1} \to A^{2n+1}$ be the projection which maps $u \in A^{2n+2M+2N+1}$ to $u_{M+N+1}^{M+N+2n+1}$; let μ_n' be the conditional PMF $\mu_{2n+2N+1}$; and let p_n be the PMF on $A^{2n+2M+2N+1}$ induced by P; that is, $p_n(u) = P[U_{-n-M-N}^{n+M+N} = u]$, $u \in A^{2n+2M+2N+1}$. Then

$$E[d_n(U_{-n}^n, V_{-n}^n)] = \sum_{u \in A^{2n+2M+2N+1}} \Big[\sum_{y \in C^{2n+2N+1}} d_n(\pi_n(u), \psi_n(y))\mu_n'(y \mid \phi_n(u)) \Big] p_n(u).$$

By Chebyshev's inequality, there exists for each n a set $A_n^* \subset A^{2n+2M+2N+1}$ such that

(a) $p_n(A_n^*) > 1 - \epsilon^2$.

(b) $u \in A_n^*$ implies that $\sum_{y \in C^{2n+2N+1}} d_n(\pi_n(u), \psi_n(y))\, \mu_n'(y \mid \phi_n(u)) < \epsilon^2$.

Applying Chebyshev's inequality again, we may find for each $u \in A_n^*$ a set $C_{nu}^* \subset C^{2n+2N+1}$ such that

(c) $\mu_n'(C_{nu}^* \mid \phi_n(u)) > 1 - \epsilon$.

(d) $y \in C_{nu}^*$ implies that $d_n(\pi_n(u), \psi_n(y)) < \epsilon$.

If $v \in A^{2n+1}$ and $\delta > 0$, let $B_n(v, \delta)$ be the ball $B_n(v, \delta) = \{u \in A^{2n+1}: d_n(u, v) < \delta\}$. We have then that $u \in A_n^*, y \in C_{nu}^*$ implies that $\psi_n(C_{nu}^*) \subset B_n(\pi_n(u), \epsilon)$. Let $b_n(\delta)$ be the cardinality of each ball $B_n(v, \delta)$. It is easily seen that $b_n(\delta) = \sum_{j=0}^{[(2n+1)\delta]} \binom{2n-1}{j}(|A| - 1)^j$, where $[\cdot]$ denotes the greatest integer function. We show that there are $k = [|\pi_n(A_n^*)|/b_n(2\epsilon)]$ points $u_1, u_2, \ldots, u_k$ in $\pi_n(A_n^*)$ such that the balls $\{B_n(u_i, \epsilon): i = 1, \ldots, k\}$ are pairwise disjoint. To see this, we proceed inductively as follows: Choose $u_1 \in \pi_n(A_n^*)$ arbitrarily. Having chosen $u_1, \ldots, u_j$, where $j < k$, so that $\{B(u_i, \epsilon): i = 1, \ldots, j\}$ are pairwise disjoint, choose u_{j+1} from the set $\pi_n(A_n^*) \setminus [\bigcup_{i=1}^j B(u_i, 2\epsilon)]$. (This set is nonempty because its cardinality is at least $|\pi_n(A_n^*)| - jb_n(2\epsilon)$, which is greater than $|\pi_n(A_n^*)| - kb_n(2\epsilon) \geq 0$.) Then $\{B(u_i, \epsilon): i = 1, \ldots, j+1\}$ are pairwise disjoint.

Now choose $u_1^*, u_2^*, \ldots, u_k^* \in A_n^*$ so that $\pi_n(u_i^*) = u_i$, $i = 1, \ldots, k$. Then $C_{nu_i^*}^* \cap C_{nu_j^*}^* = \phi$, $i \neq j$. Also, $\mu_n'(C_{nu_i^*}^* \mid \phi_n(u_i^*)) > 1 - \epsilon$ for each i. Hence, we have a $(k, 2N + 2n + 1, \epsilon)$ code for the channel $[B, C, \mu]$. For $j = 1, 2,\ldots,$ let $K(j, \epsilon)$ be the maximal k for which there exists a (k, j, ϵ) code. Then $K(2n + 2N + 1, \epsilon) \geqslant [| \pi_n(A_n^*)|/b_n(2\epsilon)]$. Since π_n is a projection it follows that $| \pi_n(A_n^*)| \geqslant | A_n^* |/| A |^{2M+2N}$. Since $P[U_{-n-M-N}^{n+M+N} \in A_n^*] > 1 - \epsilon^2$ for each n, it follows that $\liminf_{n\to\infty}(2n)^{-1} \log | A_n^* | \geqslant H(P)$. [This is a consequence of the Shannon-McMillan Theorem; see Parthasarathy (1963).] Now $b_n(2\epsilon) \leqslant [(2n + 1)\, 2\epsilon + 1]\binom{2n+1}{[(2n+1)2\epsilon]} | A |^{(2n+1)2\epsilon}$. Also,

$$\lim_{n\to\infty}(2n)^{-1} \log \binom{2n + 1}{[(2n + 1)\, 2\epsilon]} = -2\epsilon \log 2\epsilon - (1 - 2\epsilon) \log(1 - 2\epsilon)$$

(for this, see Ash, 1965, p. 115). Hence,

$$\begin{aligned}\liminf_{n\to\infty}&(2n)^{-1} \log(| \pi_n(A_n^*)|/b_n(2\epsilon))\\ &\geqslant H(P) - 2\epsilon \log 2\epsilon - (1 - 2\epsilon) \log(1 - 2\epsilon) - 2\epsilon \log | A | > 0,\end{aligned}$$

for ϵ sufficiently small. It follows that $\limsup_{j\to\infty} j^{-1}K(j, \epsilon) \geqslant H(P) - 2\epsilon \log 2\epsilon - (1 - 2\epsilon) \log(1 - 2\epsilon) - 2\epsilon \log | A |$, for ϵ sufficiently small. But $\lim_{\epsilon\to 0+} \limsup_{j\to\infty} j^{-1} \log K(j, \epsilon) = C_b$ (see Winkelbauer, 1960). Hence $C_b \geqslant H(P)$, a contradiction.

III. A Sufficient Condition for $C_s = C_b$

In this section we find a sufficient condition (Theorem 3) on the channel $[B, C, \mu]$ so that $C_s = C_b$ will follow. In the applications section which follows this section, we will see that Theorem 3 suffices to show that $C_s = C_b$ for many types of channels considered in the literature.

First, we introduce some notation. If $(\Omega, \mathscr{F}, P)$ is a probability space, and X is a simple measurable function with domain Ω, $P(X)$ denotes the random variable with domain Ω such that $P(X)(\omega) = P[X = X(\omega)]$, $\omega \in \Omega$. Let $H_P(X)$ denote the entropy of X; we have $H_P(X) = E_P[-\log P(X)]$. If in addition Y is any measurable function with domain Ω (or sequence of measurable functions), $P(X \mid Y)$ is the random variable on Ω such that $P(X \mid Y)(\omega) = P[X = X(\omega) \mid Y](\omega)$. The conditional entropy of X given Y is $H_P(X \mid Y) = E_P[-\log P(X \mid Y)]$. The mutual information $I_P(X, Y)$ is $H_P(X) - H_P(X \mid Y)$. If Z is another measurable function or sequence of measurable functions on Ω, the conditional information density $i_P(X, Y \mid Z)$ is $\log[P(X \mid Y, Z)/P(X \mid Z)]$. The conditional mutual information of (X, Y) given Z is $I_P(X, Y \mid Z) = E_P[i_P(X, Y \mid Z)] = H_P(X \mid Z) - H_P(X \mid Y, Z)$. We assume the reader is familiar with the standard properties of entropy and information which may be found in Gallager (1968).

Let $\{X_i\}_{-\infty}^{\infty}$ and $\{Y_i\}_{-\infty}^{\infty}$ be the sequences of functions on $B^\infty \times C^\infty$ such that $X_i(x, y) = x_i$, $Y_i(x, y) = y_i$, $i \in Z$, $x \in B^\infty$, $y \in C^\infty$. (The reader may have noticed here that earlier we defined $\{X_i\}$ and $\{Y_i\}$ as sequences of functions on $A^\infty \times B^\infty \times C^\infty \times A^\infty$; no possible confusion should result, as it should be clear from the context what the domain is.) Let $T_{B\times C}$ be the transformation on $B^\infty \times C^\infty$ such that $T_{B\times C}(x, y) = (T_B x, T_B y)$, $x \in B^\infty$, $y \in C^\infty$. Let $\mathscr{M}_{B\times C}$ be the sub-σ-field of $\mathscr{B}^\infty \times \mathscr{C}^\infty$ consisting of the $T_{B\times C}$-invariant sets.

LEMMA 1. *Let t be a positive integer. For every probability measure P on $\mathscr{B}^\infty \times \mathscr{C}^\infty$ stationary with respect to $T_{B\times C}$, $\{-n^{-1} \log P(X_1{}^n) + n^{-1} \sum_{i=t}^n \log P(X_i \mid X_{i+1}^n, Y_{i-t+1}^n)\}$ converges in $L^1(P)$ as $n \to \infty$ to $E_P[i_P(X_0, Y_{-t+1}^\infty \mid X_1{}^\infty) \mid \mathscr{M}_{B\times C}]$. Furthermore, there is a measurable $T_{B\times C}$-invariant function i_t: $B^\infty \times C^\infty \to [0, \log \mid B \mid]$ such that $E_P[i_P(X_0, Y_{-t+1}^\infty \mid X_1{}^\infty) \mid \mathscr{M}_{B\times C}] = i_t$ a.e. $[P]$ for every probability measure P on $\mathscr{B}^\infty \times \mathscr{C}^\infty$ stationary with respect to $T_{B\times C}$.*

Proof. $\{-n^{-1} \log P(X_1{}^n)\}$ converges to $E_P[-\log P(X_0 \mid X_1{}^\infty) \mid \mathscr{M}_{B\times C}]$ in $L^1(P)$ by the Shannon-McMillan Theorem (see Parry, 1969, Theorem 2.5). Now $\log P(X_0 \mid X_1{}^j, Y_{-t+1}^j) \to \log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty)$ as $j \to \infty$ in $L^1(P)$ (see Parry, 1969, Theorem 2.2). Also,

$$n^{-1} \sum_{j=0}^{n-t} \log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty) \cdot T_{B\times C}^{-j} \to E_P[\log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty) \mid \mathscr{M}_{B\times C}]$$

as $n \to \infty$ in $L^1(P)$, by the Mean Ergodic Theorem (see Dunford and Schwartz, 1958, p. 667). It is easily seen that the sequences

$$\left\{n^{-1} \sum_{j=0}^{n-t} \log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty) \cdot T_{B\times C}^{-j}\right\}_{n=t}^\infty$$

and

$$\left\{n^{-1} \sum_{j=0}^{n-t} \log P(X_0 \mid X_1{}^j, Y_{-t+1}^j) \cdot T_{B\times C}^{-j}\right\}_{n=t}^\infty$$

have the same limit in $L^1(P)$ by showing that the $L^1(P)$-norm of their difference goes to zero. Hence, the second sequence converges to $E_P[\log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty) \mid \mathscr{M}_{B\times C}]$ since the first one does. Since this limit is $T_{B\times C}$-invariant,

$$\left[n^{-1} \sum_{j=0}^{n-t} \log P(X_0 \mid X_1{}^j, Y_{-t+1}^j) \cdot T_{B\times C}^{-j}\right] \cdot T_{B\times C}^n$$

$$= n^{-1} \sum_{i=t}^{n} \log P(X_i \mid X_{i+1}^n, Y_{i-t+1}^n)$$

also converges in $L^1(P)$ to $E_P[\log P(X_0 \mid X_1{}^\infty, Y_{-t+1}^\infty) \mid \mathscr{M}_{B\times C}]$. Hence

$$-n^{-1} \log P(X_1{}^n) + n^{-1} \sum_{i=t}^{n} \log P(X_i \mid X_{i+1}^n, Y_{i-t+1}^n)$$

converges to

$$-E_P[\log P(X_0 \mid X_1^{\infty}) \mid \mathscr{M}_{B\times C}] + E_P[\log P(X_0 \mid X_1^{\infty}, Y_{-t+1}^{\infty}) \mid \mathscr{M}_{B\times C}]$$
$$= E_P[i_P(X_0, Y_{-t+1}^{\infty} \mid X_1^{\infty}) \mid \mathscr{M}_{B\times C}].$$

To show the existence of i_t one proceeds exactly as in the proofs of Theorem 2.6 of Parthasarathy (1961) and Lemma 3.1 of Parthasarathy (1963).

LEMMA 2. *Let $[A, P]$ be a source and let $[B, C, \mu]$ be a channel. We suppose ϕ: $A^{\infty} \to B^{\infty}$ is an invariant coder and ψ: $C^{\infty} \to A^{\infty}$ is a sliding block decoder such that $P^{\phi,\psi}[U_0 \neq V_0] < \epsilon$. Then there exists a sliding block coder ϕ': $A^{\infty} \to B^{\infty}$ such that $P^{\phi',\psi}[U_0 \neq V_0] < \epsilon$.*

Proof. Gray and Ornstein (1976) show how to do this.

THEOREM 3. *Suppose for some positive integer t both of the conditions below hold for the channel $[B, C, \mu]$. Then $C_s = C_b$, and if $[A, P]$ is any aperiodic ergodic source with $H(P) < C_b$, $[A, P]$ is transmissible over the channel with respect to sliding block coding.*

Condition 1. *If $x_1, x_2 \in B^{t-1}$, if k is a positive integer, if $y \in C^{k+t-1}$ and if $x \in B^k$, then $\mu_{k+t-1}(y \mid x_1x) = \mu_{k+t-1}(y \mid x_2x)$. (This condition is vacuous if $t = 1$.)*

Condition 2. *For any $\epsilon > 0$, there exists $R > 0$ and a source $[B, Q]$ such that $Q\mu[i_t > R] > 1 - \epsilon$.*

Proof. The method of proof we employ here is a generalization of the technique used in the Gray and Ornstein (1976) paper for the DMC. Fix an arbitrary aperiodic ergodic source $[A, P]$ with $H(P) < C_b$. If we can show that $[A, P]$ is transmissible with respect to sliding block coding, then $C_s \geqslant C_b$ will follow. Then Theorem 2 is applied to conclude $C_s = C_b$. Fix $\epsilon > 0$. To show that $[A, P]$ is transmissible with respect to sliding block coding, by Lemma 2 we need only find an invariant coder ϕ: $A^{\infty} \to B^{\infty}$ and a sliding block decoder ψ: $C^{\infty} \to A^{\infty}$ such that $P^{\phi,\psi}[U_0 \neq V_0] < 5\epsilon$. Pick Q, R so that $Q\mu[i_t > R] > 1 - \epsilon$. By Lemma 1, $\{-n^{-1} \log Q\mu(X_1^n) - n^{-1} \log |B|^{t-1} + n^{-1} \sum_{i=t}^{n} \log Q\mu(X_i \mid X_{i+1}^n, Y_{i-t+1}^n)$ converges to i_t in $L^1(Q\mu)$. Hence, for n sufficiently large,

$$Q\mu\left[Q\mu(X_1^n) < 2^{-nR} |B|^{1-t} \prod_{i=t}^{n} Q\mu(X_i \mid X_{i+1}^n, Y_{i-t+1}^n)\right] > 1 - \epsilon.$$

Choose a conditional PMF $q^0(x \mid y)$ defined for $x \in B$, $y \in C^t$, such that $q^0(x \mid y) = Q\mu(X_i = x \mid Y_{i-t+1}^i = y)$, $i \in Z$, if $Q\mu(Y_{i-t+1}^i = y) > 0$. For $s = 1, 2,\ldots$, choose a conditional PMF $q^s(x \mid x', y)$ defined for $x \in B$, $x' \in B^s$, $y \in C^{s+t}$, such that $q^s(x \mid x', y) = Q\mu(X_i = x \mid X_{i+1}^{i+s} = x', Y_{i-t+1}^{i+s} = y)$, if $Q\mu(X_{i+1}^{i+s} = x', Y_{i-t+1}^{i+s} = y) > 0$, $i \in Z$. For each n, let q_n be the PMF on B^n such that $q_n(x) =$

$Q\mu(X_1{}^n = x)$, $x \in B^n$. Let $q_n{}'$ be the probability measure defined on subsets of $B^n \times C^n$ such that $q_n{}'(E) = \sum_{x \in B^n} q_n(x)\, \mu_n(E_x \mid x)$, $E \subset B^n \times C^n$, where $E_x = \{ y \in C^n : (x, y) \in E\}$. We have then for n sufficiently large a set $E \subset B^n \times C^n$ such that $q_n{}'(E) > 1 - \epsilon$ and $(x, y) \in E$ implies

$$q_n(x) < 2^{-nR} \mid B \mid^{1-t} \prod_{i=t}^{n} q^{n-i}(x_i \mid x_{i+1}^n, y_{i-t+1}^n).$$

The following hold:

(a) For $i = 2,\dots, n$,

$$\sum_{x \in B^n} q_n(x)\, \mu_n(E_x \mid x_i{}^n, w) \leqslant 2^{-nR}, \qquad w \in B^{i-1}.$$

(b) $\sum_{x \in B^n} q_n(x)\, \mu_n(E_x \mid w) \leqslant 2^{-nR}, \qquad w \in B^n.$

(c) $\sum_{x \in B^n} q_n(x)\, \mu_n(E_x \mid x) > 1 - \epsilon.$

For example, to see that (a) holds with $i = 2$, observe that

$$\sum_{x \in B^n} q_n(x)\, \mu_n(E_x \mid x_2{}^n, w)$$

$$\leqslant 2^{-nR} \sum_{x \in B^n, y \in C^n} \mid B \mid^{1-t} \prod_{i=t}^{n} q^{n-i}(x_i \mid x_{i+1}^n, y_{i-t+1}^n) \mu_n(y \mid x_2{}^n, w).$$

The sum on the right can be seen to be one if the variables are summed over in the following order: $x_1, x_2,\dots, x_t, y_1, x_{t+1}, y_2,\dots, x_n, y_{n-t+1}, y_{n-t+2},\dots, y_n$. Keep in mind that if $j \geqslant t - 1$, then $\mu_j(y' \mid x')$ does not depend on the first $t - 1$ coodinates of x', by Condition 1.

Since $H(P) < C_b$, we have by Theorem 1 that for sufficiently large n, there is a map ϕ': $A^{n^2+1} \to B^{n^2+1}$ and a map ψ': $C^{n^2+1} \to A^{n^2+1}$ such that

(d) $\sum_{\mathbf{u} \in A^{n^2-1}} P[u \in A^\infty : u_0^{n^2} = \mathbf{u}]\, \mu_{n^2+1}((\psi')^{-1}(\mathbf{u}) \mid \phi'(\mathbf{u})) > 1 - \epsilon.$

Fix n so large that $(n^2 + 1)/(n^2 + n + 1) > 1 - \epsilon$, $n^2 2^{-nR} < \epsilon$, and (a)–(d) hold.

Since $[A, P]$ is aperiodic, we may use the strong form of Rohlin's Theorem (see Shields, 1973, p. 22) to obtain a set $F \in \mathscr{A}^\infty$ such that:

(e) $\{T^i F : i = -n, -n + 1,\dots, n^2\}$ are pairwise disjoint.

(f) $P\left(\bigcup_{i=-n}^{n^2} T^i F\right) > 1 - \epsilon.$

(g) $P[\{u \in A^\infty : u_0^{n^2} = \mathbf{u}\} \cap F] = P\{u \in A^\infty : u_0^{n^2} = \mathbf{u}\}\, P(F), \mathbf{u} \in A^{n^2+1}.$

For each fixed $v \in B^n$, we construct an invariant coder ϕ_v: $A^\infty \to B^\infty$ and a sliding block decoder ψ_v: $C^\infty \to A^\infty$ as follows:

If $u \in A^\infty$, to determine $\phi_v(u) \in B^\infty$ first find all $k \in Z$ such that $T^k u \in F$. For each such k, define $\phi_v(u)_k^{k+n^2} = \phi'(u_k^{k+n^2})$ and $\phi_v(u)_{k-n}^{k-1} = v$. Define all the other coordinates of u not accounted for in this way to be x^*, where x^* is some fixed element of B.

In order to decode a sequence $y \in C^\infty$, first determine if there is an integer i such that $0 \leqslant i \leqslant n^2$ and $y_{-n-i}^{-1-i} \in E_v$. If there is, let i' be the smallest such integer and define $\psi_v(y)_0 \in A$ to be $[\psi'(y_{-i'}^{-i'+n^2})]_{i'+1}$. If there isn't, define $\psi_v(y)_0 = u^*$, where u^* is some fixed element of A. There is a unique invariant decoder ψ_v: $C^\infty \to A^\infty$ such that $\psi_v(y)_0$ is as we have defined it for every $y \in C^\infty$. In fact, ψ_v is easily seen to be a sliding block decoder of type $n^2 + n$.

Let us write P_v to denote the measure P^{ϕ_v,ψ_v}. Then, arguing exactly as in Gray and Ornstein (1976), one obtains the following estimates:

(h) $$P_v[U_0 \neq V_0] \leqslant P_v[T^{-i}U \notin F, 0 \leqslant i \leqslant n^2] + \sum_{i=0}^{n^2} P_v[T^{-i}U \in F,\ Y_{-n-i}^{-1-i} \notin E_v] + \sum_{i=0}^{n^2} P_v[T^{-i}U \in F, \psi'(Y_{-i}^{-i+n^2}) \neq U_{-i}^{-i+n^2}] + \sum_{0 \leqslant j < i \leqslant n^2} P_v[T^{-i}U \in F,\ Y_{-n-j}^{-1-j} \in E_v].$$

(i) $$P_v[T^{-i}U \notin F, 0 \leqslant i \leqslant n^2] = 1 - \sum_{i=0}^{n^2} P(T^iF) < 1 - (1-\epsilon)(n^2+1)/(n^2+n+1) < 1-(1-\epsilon)(1-\epsilon) < 2\epsilon.$$

(j) $$\sum_{i=0}^{n^2} P_v[T^{-i}U \in F,\ Y_{-n-i}^{-1-i} \notin E_v] = \sum_{i=0}^{n^2} P(T^iF)\, \mu_n([C_n/E_v] \mid v) \leqslant 1 - \mu_n(E_v \mid v).$$

(k) $$\sum_{i=0}^{n^2} P_v[T^{-i}U \in F, \psi'(Y_{-i}^{-i+n^2}) \neq U_{-i}^{-i+n^2}] = \sum_{i=0}^{n^2} \sum_{\mathbf{u} \in A^{n^2-1}} P[T^iF \cap \{u \in A^\infty : u_{-i}^{-i+n^2} = \mathbf{u}\}] \times (1 - \mu_{n^2+1}((\psi')^{-1}(\mathbf{u}) \mid \phi'(\mathbf{u}))) < \epsilon.$$

(l) If $0 \leqslant j < i \leqslant n^2$ and $i \geqslant j + n$, then

$$P_v[T^{-i}U \in F,\ Y_{-n-j}^{-i-j} \in E_v] = \sum_{\mathbf{u} \in A^{n^2-1}} P_v[T^{-i}U \in F,\ U_{-i}^{-i+n^2} = \mathbf{u}]\, \mu_n(E_v \mid \phi'(\mathbf{u})_{i-j-n+1}^{i-j}).$$

(m) If $0 \leqslant j < i \leqslant n^2$ and $i < j + n$, then

$$P_v[T^{-i}U \in F,\ Y_{-n-j}^{-1-j} \in E_v]$$
$$= \sum_{\mathbf{u} \in A^{n^2-1}} P_v[T^{-i}U \in F,\ U_{-i}^{-i+n^2} = \mathbf{u}]\, \mu_n(E_v \mid v_{i-j-1}^{n}, \phi'(\mathbf{u})_1^{i-j}).$$

Using (a)–(c), we see that

$$\sum_{v \in B^n} q_n(v)\, P_v[U_0 \neq V_0]$$
$$\leqslant 3\epsilon + \sum_{v} q_n(v)(1 - \mu_n(E_v \mid v))$$
$$+ \sum_{0 \leqslant j < i \leqslant n^2} 2^{-Rn} P(T^i F) \leqslant 3\epsilon + \epsilon + n^2 2^{-Rn} < 5\epsilon.$$

Hence, for some choice of v, $P_v[U_0 \neq V_0] < 5\epsilon$, and the proof is complete.

IV. Applications

In this section, we show using Theorem 3 that $C_s = C_b$ for three types of channels: channels with additive random noise, ergodic channels, and averaged channels whose components are DMC's.

The Channel with Additive Random Noise

Let B be a finite abelian group. Then B^∞ is a group in the usual way: If $x = (x_i)$ and $x' = (x_i') \in B^\infty$, then define $x + x' = (x_i + x_i')$. If $[B, Q]$ is a source, define the channel $[B, B, \mu]$ with additive random noise Q as follows: $\mu(E \mid x) = Q(E - x)$, $x \in B^\infty$, $E \in \mathscr{B}^\infty$. As usual, $X = \{X_i\}$ and $Y = \{Y_i\}$ are the sequences of measurable functions mapping $B^\infty \times B^\infty$ to B such that $X_i(x, y) = x_i$, $Y_i(x, y) = y_i$, $x, y \in B^\infty$, $i \in Z$. Define $W = \{W_i\}_{-\infty}^{\infty}$ to be the sequence of projections from B^∞ to B: $W_i(x) = x_i$, $x \in B^\infty$, $i \in Z$. Let $[B, P]$ be the iid source such that $P[W_0 = x] = |B|^{-1}$, $x \in B$. The distribution of the sequence $Y - X = \{Y_i - X_i\}$ under $P\mu$ is the same as the distribution of W under Q. We have the following:

(a) If $x, y \in B^\infty$, then

$$P\mu(X_0 \mid X_1^n, Y_0^n)(x, y)$$
$$= \frac{P\mu(Y_0^n = y_0^n \mid X_0^n = x_0^n)\, P\mu(X_0^n = x_0^n)}{[\sum_{x \in B} P\mu(Y_0^n = y_0^n \mid X_0 = x, X_1^n = x_1^n)\, P\mu(X_0 = x, X_1^n = x_1^n)]}$$
$$= P\mu(Y_0^n = y_0^n \mid X_0^n = x_0^n)/P\mu(Y_1^n = y_1^n \mid X_1^n = x_1^n)$$
$$= [Q(W_0^n)/Q(W_1^n)](y - x) = Q(W_0 \mid W_1^n)(y - x).$$

This implies that:

(b) $i(X_0 \mid X_1^\infty, Y_0^\infty)(x, y) = \lim_{n\to\infty} \log[P\mu(X_0 \mid X_1^n, Y_0^n)/P\mu(X_0 \mid X_1^n)](x, y) = \log \mid B \mid + [\log Q(W_0 \mid W_1^\infty)](y - x)$.

If $F \in L^1(Q)$, it is easy to see that $E_{P\mu}[F(Y - X) \mid \mathcal{M}_{B\times B}] = E_Q[F \mid \mathcal{M}_B](Y - X)$ a.e. $[P\mu]$. Hence, taking the expected value of (b) with respect to $\mathcal{M}_{B\times B}$, we get $i_1 = \log \mid B \mid + E_Q[\log Q(W_0 \mid W_1^\infty) \mid \mathcal{M}_B](Y - X)$ a.e. $[P\mu]$. Thus ess $\inf_{P\mu} i_1 = \log \mid B \mid +$ ess $\inf_Q E_Q[\log Q(W_0 \mid W_1^\infty) \mid \mathcal{M}_B]$, which is equal to C_b by Parthasarathy (1963) or Kieffer (1974a). If $C_b > 0$, then $P\mu[i_1 > R] = 1$ for $0 < R < C_b$. Applying Theorems 2 and 3 we see that for the channel with additive random noise, $C_s = C_b$ and any aperiodic ergodic source with entropy smaller than C_b is transmissible with respect to sliding block coding.

Ergodic Channels

A channel $[B, C, \mu]$ is ergodic if for any ergodic source $[B, P]$, $P\mu$ is ergodic with respect to $T_{B\times C}$. We prove our results here for a more general class of channels called weakly ergodic channels. We define the channel $[B, C, \mu]$ to be weakly ergodic if there exists an iid source $[B, P]$ such that $P[W_0 = x] > 0$ for all $x \in B$ and $P\mu$ is ergodic.

DEFINITION. If S is a finite set, and p is a PMF on S, define $H_1(p)$ to be the entropy of p: $H_1(p) = -\sum_{s\in S} p(s) \log p(s)$. The set of PMF's on S is a convex set in the natural way.

LEMMA 3. *The entropy function H_1 is a strictly concave function of the PMF's on S. That is, if $p_1, p_2, \ldots, p_k$ are PMF's on S and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are positive numbers summing to one, then $H_1(\sum_{i=1}^k \alpha_i p_i) \geqslant \sum_{i=1}^k \alpha_i H_1(p_i)$, with equality only if $p_1 = p_2 = \cdots = p_k$.*

Lemma 3 is a well-known result. See Gallager (1968, p. 85).

LEMMA 4. *Let $[B, C, \mu]$ be a channel. Let $[B, P]$ be an iid source such that $P[W_0 = x] > 0$, $x \in B$. Let t be a positive integer. If $I_{P\mu}(X_1^t, Y_1^\infty \mid X_{t+1}^\infty) = 0$, then for any positive integer k, any $x_1, x_2 \in B^t$, any $x \in B^k$, and any $y \in C^{k+t}$, we have $\mu_{k+t}(y \mid x_1 x) = \mu_{k+t}(y \mid x_2 x)$.*

Proof. Fix the positive integer k. For $j = 1, 2, \ldots$, let p_j be the PMF on B^j such that $p_j(x) = P[W_1^j = x]$, $x \in B^j$. Since $I_{P\mu}(X_1^t, Y_1^{t+k} \mid X_{t+1}^\infty) \leqslant I_{P\mu}(X_1^t, Y_1^\infty \mid X_{t+1}^\infty)$, we have $I_{P\mu}(X_1^t, Y_1^{t+k} \mid X_{t+1}^\infty) = H(X_1^t \mid X_{t+1}^\infty) - H(X_1^t \mid Y_1^{t+k}, X_{t+1}^\infty) = 0$, where for the rest of the proof all entropies and information are calculated with respect to $P\mu$. Now $H(X_1^t \mid Y_1^{t+k}, X_{t+1}^{t+k}) \geqslant H(X_1^t \mid Y_1^{t+k}, X_{t+1}^\infty)$. Also, $H(X_1^t \mid X_{t+1}^\infty) = H(X_1^t \mid X_{t+1}^{t+k})$, since the distribution of $\{X_i\}$ under $P\mu$ is the same as that of $\{W_i\}$ under P, and $\{W_i\}$ is an independent sequence relative to P. Hence $I(X_1^t, Y_1^{t+k} \mid X_{t+1}^{t+k}) = H(X_1^t \mid X_{t+1}^{t+k}) - H(X_1^t \mid Y_1^{t+k}, X_{t+1}^{t+k}) = 0$. We have:

$$
\begin{aligned}
\text{(a)}\quad I(X_1^t, Y_1^{t+k} \mid X_{t+1}^{t+k}) &= H(Y_1^{t+k} \mid X_{t+1}^{t+k}) - H(Y_1^{t+k} \mid X_1^{t+k}) \\
&= \sum_{x\in B^{t+k}, y\in C^{t+k}} p_{t+k}(x)\, \mu_{t+k}(y \mid x) \\
&\quad \times \log \left[\mu_{t+k}(y \mid x) \Big/ \sum_{x'\in B^t} p_t(x')\mu_{t+k}(y \mid x' x_{t+1}^{t+k})\right] \\
&= \sum_{\bar{x}\in B^k} p_k(\bar{x}) \left[H_1\left(\sum_{x'\in B^t} p_t(x')\, \mu_{t+k}(\cdot \mid x'\bar{x})\right)\right. \\
&\quad \left. - \sum_{x'\in B^t} p_t(x')\, H_1(\mu_{t+k}(\cdot \mid x'\bar{x}))\right] = 0.
\end{aligned}
$$

Since in this last sum all $p_k(\bar{x}) > 0$ and all $p_t(x') > 0$, we apply Lemma 3 to conclude that for any $\bar{x} \in B^k$, the PMF's $\{\mu_{t+k}(\cdot \mid x'\bar{x}) : x' \in B^t\}$ coincide.

Now suppose $[B, C, \mu]$ is a weakly ergodic channel. Suppose $C_b > 0$. Let $[B, P]$ be an iid source such that $P[W_0 = x] > 0$, $x \in B$, and $P\mu$ is ergodic. Then for some positive integer t we have $I_{P\mu}(X_1^t, Y_1^\infty \mid X_{t+1}^\infty) > 0$ and $I_{P\mu}(X_1^i, Y_1^\infty \mid X_{i+1}^\infty) = 0$, $1 \leqslant i < t$. (Otherwise, by Lemma 4, the measures $\{\mu(\cdot \mid x) : x \in B^\infty\}$ are identical and so $C_b = 0$.) Lemma 4 implies that Condition 1 of Theorem 3 holds. Now we show that Condition 2 holds. In the following let all information be calculated with respect to $P\mu$. Since $P\mu$ is ergodic, we have $i_t = I(X_0, Y_{-t+1}^\infty \mid X_1^\infty)$ a.e. $[P\mu]$. Now,

$$
\begin{aligned}
I(X_1^t, Y_1^\infty \mid X_{t+1}^\infty) &= \sum_{i=1}^t I(X_i, Y_1^\infty \mid X_{i+1}^\infty) = \sum_{i=1}^t I(X_0, Y_{1-i}^\infty \mid X_1^\infty) \\
&\leqslant \sum_{i=1}^t I(X_0, Y_{1-t}^\infty \mid X_1^\infty) = tI(X_0, Y_{1-t}^\infty \mid X_1^\infty).
\end{aligned}
$$

Hence $I(X_0, Y_{1-t}^\infty \mid X_1^\infty) > 0$. Thus, $P\mu[i_t > R] = 1$ if $0 < R < I(X_0, Y_{1-t}^\infty \mid X_1^\infty)$ and Condition 2 holds. We conclude that for a weakly ergodic channel (and therefore for an ergodic channel) $C_b = C_s$ and any aperiodic ergodic source with entropy less than C_b is transmissible with respect to sliding block coding.

Averaged Channels Whose Components Are DMC's

Let $(S, \mathscr{S}, \alpha)$ be a probability space. For each $s \in S$, let $[B, C, \mu^s]$ be a DMC. We suppose that for $E \in \mathscr{C}^\infty$ and $x \in B^\infty$, the map $s \to \mu^s(E \mid x)$ with domain S is measurable. Let $[B, C, \mu]$ be the averaged channel whose components are DMC's defined as follows: $\mu(E \mid x) = \int_S \mu^s(E \mid x)\, d\alpha(s)$, $x \in B^\infty$, $E \in \mathscr{C}^\infty$. If

$[B, P]$ is an ergodic source it is not hard to show that $P\mu(F) = \int_S P\mu^s(F)\, d\alpha(s)$, $F \in \mathscr{B}^\infty \times \mathscr{C}^\infty$. Hence, $P\mu[i_1 > R] = \int_S P\mu^s[i_1 > R]\, d\alpha(s)$. Since for each s, $P\mu^s$ is ergodic, we have that $i_1 = I_{P\mu^s}(X_0, Y_0^\infty \mid X_1^\infty)$ a.e. $[P\mu^s]$. If $[B, P]$ is an iid source then $\{(X_i, Y_i)\}$ is an independent sequence relative to $P\mu^s$, from which it follows that $I_{P\mu^s}(X_0, Y_0^\infty \mid X_1^\infty) = I_{P\mu^s}(X_0, Y_0)$. Hence, for each s, $i_1 = I_{P\mu^s}(X_0, Y_0)$ a.e.$[P\mu^s]$, if $[B, P]$ is an iid source. Now for this type of averaged channel, $C_b = \lim_{\lambda\to 0+} \sup_{P\in\mathscr{P}} \sup_{\{E\in\mathscr{S}:\alpha(E)>1-\lambda\}} \inf_{s\in E} I_{P\mu^s}(X_0, Y_0)$, where $\mathscr{P} = \{P: [B, P]$ is an iid source$\}$ (see Ahlswede, 1968, or Kieffer, 1974a, for this formula). Assume $C_b > 0$. Then given $\epsilon > 0$ and $0 < R < C_b$, there exists $P \in \mathscr{P}$ and $E \in \mathscr{S}$ such that $I_{P\mu^s}(X_0, Y_0) > R$ for $s \in E$ and $\alpha(E) > 1 - \epsilon$. Thus, $P\mu[i_1 > R] = \int_S P\mu^s[i_1 > R]\, d\alpha(s) \geqslant \int_E P\mu^s[I_{P\mu^s}(X_0, Y_0) > R]\, d\alpha(s) = \alpha(E) > 1 - \epsilon$. We now apply Theorem 3 to conclude that for an averaged channel whose components are DMC's, $C_s = C_b$ holds and any aperiodic ergodic source with entropy less then C_b is transmissible with respect to sliding block coding.

V. Some Remarks on Periodic Sources

For the types of channels considered in the preceding applications, we are able to transmit over the channel with respect to sliding block coding any aperiodic ergodic source with entropy less than C_s. A natural question to ask is whether in general the same result holds for periodic ergodic sources. We show that this is not true. To show this, since a periodic ergodic source has entropy zero, we find, for each type of channel previously considered, a channel of that type with $C_s > 0$ and a periodic ergodic source which is not transmissible with respect to sliding block coding.

Definition. If $[A, P]$ is a periodic ergodic source, then for some positive integer n, there is a periodic sequence $u^* \in A^\infty$ of period n such that $P\{T_A{}^i u^*\} = n^{-1}, i = 0, 1,..., n - 1$. We call n the period of the periodic ergodic source $[A, P]$.

Theorem 4. *Let $[B, C, \mu]$ be a channel and let $[A, P]$ be a periodic ergodic source of period n. Then $[A, P]$ is transmissible over $[B, C, \mu]$ with respect to sliding block coding if and only if there exists a periodic sequence $x^* \in B^\infty$ of period n such that the n measures $\{\mu(\cdot \mid T_B{}^i x^*): i = 0, 1,..., n - 1\}$ are mutually singular.*

Proof. Suppose $x^* \in B^\infty$ is periodic with period n and $\{\mu(\cdot \mid T_B{}^i x^*): i = 0, 1,..., n - 1\}$ are mutually singular. Suppose $[A, P]$ is a periodic ergodic source of period n. Let $u^* \in A^\infty$ be periodic of period n such that $P\{T_A{}^i u^*\} = n^{-1}$, $i = 0, 1,..., n - 1$. Given $\epsilon > 0$, we may find a positive integer k and a partition $\{E_0, E_1, ..., E_{n-1}\}$ of C^k such that $\mu_k(E_i \mid [T_B{}^i x^*]_1^k) > 1 - \epsilon$, $i = 0,..., n - 1$. Define an invariant coder ϕ: $A^\infty \to B^\infty$ as follows: Define $\phi(T_A{}^i u^*) =$

$T_B{}^i x^*$, $i = 0,..., n - 1$. For all other $u \in A^\infty$, define $\phi(u) = x'$, where x' is a fixed constant sequence in B^∞. Define ψ: $C^\infty \to A^\infty$ to be the unique sliding block decoder such that $\psi(y)_0 = [T_A{}^i u^*]_0$ if $y_1{}^k \in E_i$, $i = 0,..., n - 1$. Then $P^{\phi,\psi}[U_0 = V_0] = n^{-1} \sum_{i=0}^{n-1} \mu_k(E_i \mid \phi(T_A{}^i u^*)_1^k) > 1 - \epsilon$. Thus $[A, P]$ is transmissible with respect to sliding block coding.

Conversely, suppose $[A, P]$ is transmissible with respect to sliding block coding. For each $\epsilon > 0$, choose invariant ϕ_ϵ: $A^\infty \to B^\infty$ and ψ_ϵ: $C^\infty \to A^\infty$ such that $P^{\phi_\epsilon, \psi_\epsilon}[U_0^{n-1} = V_0^{n-1}] > 1 - \epsilon/n$. We have $P^{\phi_\epsilon, \psi_\epsilon}[U_0^{n-1} = V_0^{n-1}] = n^{-1} \sum_{i=0}^{n-1} \mu(\{y \in C^\infty: \psi_\epsilon(y)_0^{n-1} = [T_A{}^i u^*]_0^{n-1}\} \mid \phi_\epsilon(T_A{}^i u^*))$. Since the n sequences $\{[T_A{}^i u^*]_0^{n-1}: i = 0,..., n - 1\}$ are distinct, the sets $\{E_i^\epsilon: i = 0,..., n - 1\}$ are pairwise disjoint, where $E_i^\epsilon = \{y \in C^\infty: \psi_\epsilon(y)_0^{n-1} = [T_A{}^i u^*]_0^{n-1}\}$. Let $x_\epsilon^* \in B^\infty$ be the sequence $\phi_\epsilon(u^*)$. We have:

(a) $\mu(E_i^\epsilon \mid T_B{}^i x_\epsilon^*) > 1 - \epsilon$, $i = 0, 1,..., n - 1$.

If $\epsilon < \frac{1}{2}$, we see from (a) that the sequences $\{T_B{}^i x_\epsilon^*: i = 0,..., n - 1\}$ are distinct. Since $T_A{}^n u^* = u^*$ and ϕ_ϵ is invariant, we must have $T_B{}^n x_\epsilon^* = x_\epsilon^*$. Hence for $\epsilon < \frac{1}{2}$, x_ϵ^* has period n. Since there are only finitely many periodic sequences in B^∞ of period n, x_ϵ^* must remain constant as ϵ approaches zero through some sequence $\{\epsilon_n\}$. Let x^* be this constant value of x_ϵ^* through $\{\epsilon_n\}$. Then (a) implies that $\{\mu(\cdot \mid T_B{}^i x^*): i = 0,..., n - 1\}$ are mutually singular.

COROLLARY. *Let $[B, C, \mu]$ be a channel where $B = \{0, 1\}$. Let $x \in B^\infty$ be the periodic sequence with period two such that $x_0 = 0$, $x_1 = 1$. Let $[B, P]$ be the periodic ergodic source such that $P\{x\} = P\{T_B x\} = \frac{1}{2}$. Then $[B, P]$ is transmissible over the channel $[B, C, \mu]$ with respect to sliding block coding if and only if $\mu(\cdot \mid x)$ and $\mu(\cdot \mid T_B x)$ are mutually singular.*

Proof. This follows from Theorem 4 because x and $T_B x$ are the only periodic sequences in B^∞ with period 2.

EXAMPLE. Using the above corollary, we construct an example of an ergodic channel with additive random noise which has $C_s > 0$, but some periodic ergodic source is not transmissible with respect to sliding block coding.

Fix $B = C = \{0, 1\}$, where we make B into a group using addition modulo 2.

LEMMA 5. *Let $[B, Q]$ be a source such that Q is mixing with respect to T_B. Let $[B, B, \mu]$ be the channel with additive random noise such that $\mu(E \mid x) = Q(E - x)$, $E \in \mathscr{B}^\infty$, $x \in B^\infty$. Then $[B, B, \mu]$ is an ergodic channel.*

Proof. Let $[B, P]$ be a fixed ergodic source. We show that $P\mu$ is ergodic with respect to $T_{B \times B}$. There exists a probability space $(\Omega, \mathscr{F}, \lambda)$ and two sequences of B-valued random variables $\{X_i\}_{-\infty}^\infty$ and $\{Z_i\}_{-\infty}^\infty$ such that:

(a) The sequence $\{X_i\}$ is independent of the sequence $\{Z_i\}$.

(b) The distribution of $\{X_i\}$ is P.

(c) The distribution of $\{Z_i\}$ is Q.

It is easily seen that the sequence $\{(X_i, X_i + Z_i)\}$ has distribution $P\mu$. Hence showing that $P\mu$ is ergodic with respect to $T_{B\times B}$ is equivalent to showing that the sequence $\{(X_i, X_i + Z_i)\}$ is ergodic with respect to the measure λ. But $\{(X_i, X_i + Z_i)\}$ will be ergodic if $\{(X_i, Z_i)\}$ is ergodic. To show that $\{(X_i, Z_i)\}$ is ergodic, let j, k be arbitrary positive integers. Let $x, z \in B^j$ and $x', z' \in B^k$ be arbitrary. If $\alpha_n = \lambda[X_1{}^j = x, Z_1{}^j = z, X_{n+1}^{n+k} = x', Z_{n+1}^{n+k} = z']$, then for n sufficiently large $\alpha_n = \beta_n\gamma_n$, where $\beta_n = \lambda[X_1{}^j = x, X_{n+1}^{n+k} = x']$ and $\gamma_n = \lambda[Z_1{}^j = z, Z_{n+1}^{n+k} = z']$. $\{(X_i, Z_i)\}$ will be ergodic if we show that α_n is Cesaro-summable to $\lambda[X_1{}^j = x, Z_1{}^j = z]\,\lambda[X_1{}^k = x', Z_1{}^k = z'] = \beta\gamma$, where $\beta = \lambda[X_1{}^j = x]\,\lambda[X_1{}^k = x']$, $\gamma = \lambda[Z_1{}^j = z]\,\lambda[Z_1{}^k = z']$. Now β_n is Cesaro-summable to β because $\{X_i\}$ is an ergodic sequence. γ_n converges to γ because $\{Z_i\}$ is a mixing sequence. It follows that $\beta_n\gamma_n$ is Cesaro-summable to $\beta\gamma$. To see this, fix $\epsilon > 0$. Choose m so large that $n \geqslant m$ implies that $|\gamma_n - \gamma| < \epsilon$. Then

$$\left| s^{-1}\sum_{n=1}^{s} \gamma_n\beta_n - s^{-1}\sum_{n=1}^{s} \gamma\beta_n \right|$$

$$\leqslant 2m/s + \left| s^{-1}\sum_{n=m}^{s} \gamma_n\beta_n - s^{-1}\sum_{n=m}^{s} \gamma\beta_n \right| \leqslant 2m/s + \epsilon.$$

Hence, $\limsup_{s\to\infty} | s^{-1}\sum_{n=1}^{s}\gamma_n\beta_n - s^{-1}\sum_{n=1}^{s}\gamma\beta_n | \leqslant \epsilon$. Since ϵ is arbitrary, $\lim_{s\to\infty} s^{-1}\sum_{n=1}^{s}\gamma_n\beta_n = \lim_{s\to\infty} s^{-1}\sum_{n=1}^{s}\gamma\beta_n = \gamma\beta$.

Suppose we can construct a mixing stationary sequence $Z = \{Z_i\}$ of B-valued random variables on a probability space $(\Omega, \mathscr{F}, \lambda)$ such that $H(Z) = \lim_{n\to\infty} (n+1)^{-1}H(Z_0{}^n) < \log 2$ and

$$\begin{aligned} \lambda[Z_m &= i_m, Z_{m+1} = i_{m+1}, \dots, Z_n = i_n] \\ &= \lambda[Z_m = i_m + 1, Z_{m+1} = i_{m+1} + 1, \dots, Z_n = i_n + 1], \\ &\quad m, n \in Z,\ m \leqslant n, \text{ and } i_m, i_{m+1}, \dots, i_n \in B. \end{aligned} \tag{2}$$

Then if the source $[B, Q]$ is such that Q is the distribution of Z, and $[B, B, \mu]$ is the corresponding ergodic channel with additive random noise, then $C_s = C_b = \log 2 - H(Q) = \log 2 - H(Z) > 0$ (see Parthasarathy, 1963, or Kieffer, 1974a). Let x be the periodic sequence in B^∞ with period 2 such that $x_0 = 0$, $x_1 = 1$. Then Eq. (2) implies that $\mu(\cdot \mid x) = \mu(\cdot \mid T_B x)$. The corollary to Theorem 4 would then imply that if $[B, P]$ is the periodic ergodic source such that $P\{x\} = P\{T_B x\} = \frac{1}{2}$, then $[B, P]$ is not transmissible over $[B, B, \mu]$ with respect to sliding block coding.

The following lemma will allow us to find such a sequence $\{Z_i\}$.

LEMMA 6. *Let S, T be commuting automorphisms of a probability space $(\Omega, \mathscr{F}, \lambda)$. Let T be mixing. Let $\mathscr{P} = \{A_0, A_1\}$ be a measurable partition of Ω such that $S^{-1}A_0 = A_1$ and $S^{-1}A_1 = A_0$. Define $Z_0: \Omega \to B$ as follows: $Z_0(\omega) = 0$ if $\omega \in A_0$; $Z_0(\omega) = 1$ if $\omega \in A_1$. Then the sequence $\{Z_i\}_{-\infty}^{\infty}$ where $Z_i = Z_0 \cdot T^i$, $i \in Z$, is stationary and mixing and Eq.* (2) *holds.*

Proof. The fact that $\{Z_i\}$ is stationary and mixing follows because T is measure-preserving and mixing. Also

$$\begin{aligned}
\lambda[Z_m = i_m, \ldots, Z_n = i_n] &= \lambda[T^{-m}A_{i_m} \cap \cdots \cap T^{-n}A_{i_n}] \\
&= \lambda[S^{-1}T^{-m}A_{i_m} \cap \cdots \cap S^{-1}T^{-n}A_{i_n}] \\
&= \lambda[T^{-m}(S^{-1}A_{i_m}) \cap \cdots \cap T^{-n}(S^{-1}A_{i_n})] \\
&= \lambda[T^{-m}(A_{i_m+1}) \cap \cdots \cap T^{-n}(A_{i_n+1})] \\
&= \lambda[Z_m = i_m + 1, \ldots, Z_n = i_n + 1].
\end{aligned}$$

Now we construct $\{Z_i\}$. In Lemma 6, take the measurable space $(\Omega, \mathscr{F})$ to be $(B^\infty, \mathscr{B}^\infty)$. For each $i \in Z$, let $W_i: B^\infty \to B$ be the projection $W_i(x) = x_i$, $x \in B^\infty$. Then $(B^\infty, \mathscr{B}^\infty, \lambda)$ is a probability space where $[B, \lambda]$ is the iid source such that $P[W_0 = 0] = P[W_0 = 1] = \frac{1}{2}$. In Lemma 6, we take T to be the shift T_B, which is mixing and stationary with respect to λ. To define S let $x^* \in B^\infty$ be the constant sequence such that $x_i^* = 1$, for all i. Define $S: B^\infty \to B^\infty$ so that $S(x) = x + x^*$, $x \in B^\infty$. For the partition $\mathscr{P} = \{A_0, A_1\}$ take $A_0 = \{(W_0, W_1, W_2) \in \{(0, 0, 0), (0, 1, 0), (1, 1, 0), (0, 1, 1)\}\}$. If $\{Z_i\}$ is constructed as in Lemma 6, then $H(Z_0, Z_1)/2 < \log 2$ because the partition $\mathscr{P} \cap T^{-1}\mathscr{P}$ contains four elements one of which is $\{(W_0, W_1, W_2, W_3) \in \{(0, 0, 0, 0), (0, 1, 1, 0)\}$ which has λ probability less than $\frac{1}{4}$. Hence $H(Z) < \log 2$.

ANOTHER EXAMPLE. We now construct an example of an averaged channel whose components are DMC's such that $C_s > 0$ but there exists a periodic ergodic source not transmissible with respect to sliding block coding.

Again, let $B = C = \{0, 1\}$. Let $\mu^1(\cdot \mid \cdot): B \times B \to [0, 1]$ be a conditional PMF. Define $\phi: B \to B$ so that $\phi(0) = 1, \phi(1) = 0$. Let $\mu^2(\cdot \mid \cdot): B \times B \to [0, 1]$ be the conditional PMF such that $\mu^2(j \mid i) = \mu^1(j \mid \phi(i))$, $i, j \in B$. Let $[B, B, \mu^1]$ be the DMC such that $\mu^1(\cdot \mid x) = \prod_{i=-\infty}^{\infty} \mu^1(\cdot \mid x_i)$. Let $[B, B, \mu^2]$ be the DMC such that $\mu^2(\cdot \mid x) = \prod_{i=-\infty}^{\infty} \mu^2(\cdot \mid x_i)$. Let $[B, B, \mu]$ be the averaged channel such that $\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$. Pick μ^1 so that $\mu^1(1 \mid 0) \neq \mu^1(1 \mid 1)$. Then if C_b^1 and C_b^2 are the block coding capacities of $[B, B, \mu^1]$ and $[B, B, \mu^2]$, respectively, we have $C_b^1 = C_b^2 > 0$, and hence C_b, the block coding capacity of $[B, B, \mu]$, is positive, since by the Nedoma lower bound, $C_b \geqslant [(C_b^1)^{-1} + (C_b^2)^{-1}]^{-1} > 0$ (see Kieffer, 1974a, p. 388.) Let $x \in B^\infty$ be the periodic sequence of period 2 such that $x_0 = 0$, $x_1 = 1$. Then it is easily seen that $\mu^1(\cdot \mid x) = \mu^2(\cdot \mid T_B x)$ and

$\mu^1(\cdot \mid T_B x) = \mu^2(\cdot \mid x)$. Hence, $\mu(\cdot \mid x) = \mu(\cdot \mid T_B x)$ and by the corollary to Theorem 4 the periodic ergodic source $[P, B]$ such that $P\{x\} = P\{T_B x\} = \frac{1}{2}$ is not transmissible over $[B, B, \mu]$ with respect to sliding block coding and yet $C_s = C_b > 0$.

Transmission of Periodic Sources for the DMC

The failure of the previous examples to transmit certain periodic ergodic sources does not occur for the DMC.

THEOREM 5. *Let $[B, C, \mu]$ be a DMC with $C_b > 0$. Then every periodic ergodic source is transmissible over $[B, C, \mu]$ with respect to sliding block coding.*

Proof. By Theorem 4, all we need to do is show that for each $n \geqslant 2$ there is a periodic sequence x in B^∞ of period n such that $\{\mu(\cdot \mid T_B{}^i x): i = 0, 1,..., n-1\}$ are mutually singular. Since $C_b > 0$ we may choose $b, b' \in B$ such that $\mu_1(\cdot \mid b) \neq \mu_1(\cdot \mid b')$. Fix $n \geqslant 2$. Let $x \in B^\infty$ be periodic of period n and have all of its entries drawn from the set $\{b, b'\}$. Consider the DMC $[B', C', \mu']$ such that $B' = B^n$, $C' = C^n$, and $\mu_1' = \mu_n$. For $i = 0, 1,..., n-1$, let x_i be the sequence in $(B')^\infty$ which has the constant value $[T_B{}^i x]_0^{n-1}$. Let $[B', P_i]$ be the ergodic periodic source which assigns probability 1 to x_i. Since $[B', C', \mu']$ is an ergodic channel, $P_i\mu'$ induces in $(\mathscr{C}')^\infty$ a measure ergodic and stationary with respect to the shift on $(C')^\infty$. But the measure induced by $P_i\mu'$ on $(\mathscr{C}')^\infty$ is $\mu'(\cdot \mid x_i)$. Hence, if $0 \leqslant i < j \leqslant n-1$, then either $\mu'(\cdot \mid x_i)$ and $\mu'(\cdot \mid x_j)$ are mutually singular or equal. Suppose $\mu'(\cdot \mid x_i) = \mu'(\cdot \mid x_j)$. Then $\mu_n(\cdot \mid [T_B{}^i x]_0^{n-1}) = \mu_n(\cdot \mid [T_B{}^j x]_0^{n-1})$. Since $[T_B{}^i x]_0^{n-1} \neq [T_B{}^j x]_0^{n-1}$, then for some s satisfying $0 \leqslant s \leqslant n-1$ we have $[T_B{}^i x]_s \neq [T_B{}^j x]_s$. Hence either $[T_B{}^i x]_s = b$ and $[T_B{}^j x]_s = b'$ or $[T_B{}^i x]_s = b'$ and $[T_B{}^j x]_s = b$. In either case, we conclude $\mu_1(\cdot \mid b) = \mu_1(\cdot \mid b')$, a contradiction. Thus $\{\mu'(\cdot \mid x_i): i = 0, 1,..., n-1\}$ are mutually singular. But $\mu'(\cdot \mid x_i) = \mu(\cdot \mid T_B{}^i x)$.

RECEIVED: March 19, 1976; REVISED: October 22, 1976

REFERENCES

AHLSWEDE, R. (1968), The weak capacity of averaged channels, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **11**, 61–73.

ASH, R. (1965), "Information Theory," Interscience–Wiley, New York.

DUNFORD, N., AND SCHWARTZ, J. T. (1958), "Linear Operators Part I," Interscience–Wiley, New York.

GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," John Wiley, New York.

GRAY, R. M. (1975), Sliding-block source coding, *IEEE Trans. Information Theory* **21**, 357–368.

GRAY, R. M., AND ORNSTEIN, D. S. (1976), Sliding-block joint source/noisy-channel coding theorems, *IEEE Trans. Information Theory* **22**, 682–690.

KIEFFER, J. C. (1974a), A general formula for the capacity of stationary nonanticipatory channels, *Inform. Contr.* **26**, 381–391.

KIEFFER, J. C. (1974b), On the approximation of stationary measures by periodic and ergodic measures, *Ann. Probability* **2**, 530–534.

NEDOMA, J. (1957), The capacity of a discrete channel, *in* "Transactions of First Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes, Prague," pp. 143–182.

PARRY, W. (1969), "Entropy and Generators in Ergodic Theory," W. A. Benjamin, New York.

PARTHASARATHY, K. R. (1961), On the integral representation of the rate of transmission of a stationary channel, *Illinois J. Math.* **5**, 299–305.

PARTHASARATHY, K. R. (1963), Effective entropy rate and transmission of information through channels with additive random noise, *Sankhyā Ser. A* **A25**, 75–84.

SHIELDS, P. (1973), "The Theory of Bernoulli Shifts," University of Chicago Press, Chicago.

WINKELBAUER, K. (1960), Communication channels with finite past history, *in* "Transactions of the Second Prague Conference on Information Theory, Prague," pp. 685–831.